



TITLE:

Studies on Positive and Horn Boolean Functions with Applications to Data Analysis(Abstract_要旨)

AUTHOR(S):

Makino, Kazuhisa

CITATION:

Makino, Kazuhisa. Studies on Positive and Horn Boolean Functions with Applications to Data Analysis. 京都大学, 1997, 博士(工学)

ISSUE DATE:

1997-03-24

URL:

<http://hdl.handle.net/2433/202286>

RIGHT:

氏 名	まきのかずひさ 牧 野 和 久
学位(専攻分野)	博 士 (工 学)
学 位 記 番 号	工 博 第 1587 号
学位授与の日付	平 成 9 年 3 月 24 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研究科・専攻	工 学 研 究 科 数 理 工 学 専 攻
学位論文題目	Studies on Positive and Horn Boolean Functions with Applications to Data Analysis (正およびホーンブール関数とそのデータ解析への応用に関する研究)
論文調査委員	(主 査) 教 授 茨 木 俊 秀 教 授 長 谷 川 利 治 教 授 石 田 亨

論 文 内 容 の 要 旨

本論文では、ある事象を引き起こす例(正例)のデータ集合 $T \subseteq \{0, 1\}^n$ と、引き起こさない例(負例)のデータ集合 $F \subseteq \{0, 1\}^n$ の組である部分定義論理関数 (T, F) が与えられたとき、なぜその事象が生起したと判断されるのかを説明するブール関数(拡張) f を求める問題を考察している。例えば、乳がんのデータにおいては、 T が悪性腫瘍、 F が良性腫瘍の患者のデータ集合を示し、それぞれのデータは細胞の大きさや形状の均一性、細胞核、核仁、そして有糸分裂などを示す属性に対する値からなっている。実際の医療現場では、簡単な検査で調べることができる上述の属性値を用い、腫瘍が良性か悪性かを判断し、対処する事を望んでいる。本論文では、このような判断の基となる拡張 f を現在までに得られた (T, F) から構成する問題をブール関数(とりわけ、正及びホーンブール関数)の立場から考察し、その結果をまとめたものであり、11章で構成されている。

第1章は序論であり、本論文で扱う知識獲得問題の研究状況、及び、学習理論や、データベース理論におけるデータ発掘、知識発見と呼ばれる分野との関連について述べるとともに、本研究の意義、重要性、目的および論文の構成を概説している。

第2章は、本論文で考察する問題の定式化、第3章以降で用いられる論理関数の諸定義、その基礎的な性質等について述べている。特に、正関数とホーン関数の諸性質を詳しく述べている。

第3章では、拡張 f の満たすべき構造があらかじめ指定されている場合(例えば、正、ホーン、 k -DNF、2単調、自己双対、read-once、分解可能関数など)、与えられたデータ集合 (T, F) がそのような拡張を持つかどうかを決定する問題、さらに、 (T, F) が指定された構造の拡張を持たないとき、エラー数最小の拡張を求める問題を考察している。さらに、それらの問題に対し、多項式時間で可能か、あるいはNP困難であるかを示している。

第4章では、分解可能関数の中で、特に正あるいはホーン分解可能関数に対して、第3章で考察した問題をさらに詳しく考察している。

第5章では、ホーン関数による拡張に限定し、ホーン拡張の中で最大、最小、最短であるものを求める問題を考えている。とくに、最大ホーン拡張を求める問題が人工知能分野のホーンモデル理論における問題と対応することをのべ、さらに、最小、最短のホーン拡張を求めることはNP困難であるが、極小のホーン拡張については多項式時間で求められることを示している。

第6章では、現実のデータは誤りを含んでいる場合が多いことを考慮し、誤りに強い部分を記述する関数（内核関数、外核関数）を導入し、その構造を明らかにしている。

第7章では、現実には得られるデータ集合が完全でなく、属性の値のいくつかが分からない（すなわち、missing bitsを含む）場合を考察している。例えば、上記の腫瘍の例では、ある属性に関するテストを時間やコストの制約によって実施しなかった場合がこれに相当する。まず、missing bitsの扱い方の違いによる三種類の拡張（ロバスト、コンシステント、最大ロバスト）を定義し、その性質を調べてたのち、それらの拡張を求めることが多項式時間で可能か、あるいはNP困難であることを示している。

第8章では、得られた拡張の安定性を調べるため、内包関数と外包関数を導入し、それらの基本的な性質を明らかにしている。さらに2単調な論理関数においては、これらの関数が多項式時間で構築可能であることを示している。

第9章では、知識 f を正関数に限定し、その同定問題を扱っている。より正確に言うと、 f の獲得過程において帰属性質問（あるベクトル v を与え、 $f(v)=1$ か $f(v)=0$ かを問う）を許し、 f の極小真ベクトル集合と極大偽ベクトル集合の両方を決定する問題である。この同定問題に多項式オーダーの計算量の解法が存在するかどうかは、よく知られた未解決問題である。この問題に対して、最大潜伏度という新しい概念を導入し、正関数のいくつかの部分クラスについて、この最大潜伏度を用いれば、効率のよい多項式時間同定アルゴリズムが設計できることを示している。これは、長年の未解決問題に対して、部分的ではあるが、肯定的な解決を与えたものである。

第10章では、第9章で用いられた最大潜伏度の概念を2単調なクラスに応用し、従来のアルゴリズムより簡単でかつ（現在のところ）最も高速な同定アルゴリズムを設計している。

第11章では、本論文で得られた結果を要約し、今後の課題について述べている。

論文審査の結果の要旨

ある事象を引き起こす例（正例）のデータ集合と、引き起こさない例（負例）のデータ集合の組が与えられたとき、なぜその事象が生起したと判断されるのかの説明をブール関数によって記述する問題は、人工知能分野における知識獲得の一形式であり、学習理論や、データベース理論におけるデータ発掘・知識発見の分野とも関連している。本論文は、この問題についての基礎研究をまとめたものであり、得られた主要な成果は以下の通りである。

1. 与えられたデータを説明するブール関数の存在を判定する問題を、ブール関数が正、ホーンあるいはその他の特殊な構造を持つ場合について考察し、その基本的な性質を明らかにすると共に、計算の複雑さを明らかにした。

2. 現実には得られるデータは、属性値の一部が分からなかったり誤りを含んでいる場合が多いことを考

慮し、上の議論をそれらの場合に拡張し、計算の複雑さを解明した。

3. 得られた知識の安定性を示すため、内包関数と外包関数を導入し、それらの基本的な性質を明らかにすると共に、2 単調なクラスにおいては、これらの関数が多項式時間で構築可能であることを示した。

4. 知識を正関数に限定し、その同定問題に対して、最大潜伏度という概念を導入し、正関数のいくつかの部分クラスについて、この最大潜伏度を用いれば、効率のよい多項式時間同定アルゴリズムが設計できることを示した。さらに、この結果を2 単調なクラスに応用し、(現在のところ)最も高速な同定アルゴリズムの設計に成功した。

このように、本論文は、従来から経験的に行われてきた知識獲得問題のいくつかの側面を、主にブール関数と計算の複雑さの観点から解明したものであり、学術上、實際上寄与するところが少なくない。よって、本論文は博士(工学)の学位論文として価値あるものと認める。

また平成9年1月20日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。